Human imprinted retrogenes exhibit non-canonical imprint chromatin signatures and reside in non-imprinted host genes

Published online 7 February 2011 *Nucleic Acids Research, 2011, Vol. 39, No. 11* ***4577–4586***
doi:10.1093/nar/gkq1230

Human imprinted retrogenes exhibit non-canonical imprint chromatin signatures and reside in non-imprinted host genes

David Monk[1,*], Philippe Arnaud[2], Jennifer M. Frost[3,4], Andrew J. Wood[4], Michael Cowley[4], Alejandro Martin-Trujillo[1], Amy Guillaumet-Adkins[1], Isabel Iglesias Platas[5], Cristina Camprubi[1], Deborah Bourc'his[6], Robert Feil[2], Gudrun E. Moore[3] and Rebecca J. Oakey[4]

[1]Imprinting and Cancer Group, Cancer Epigenetics and Biology Program, Bellvitge Institute for Biomedical Research (IDIBELL), L'Hospitalet de Llobregat, 08907, Barcelona, Spain, [2]Institute of Molecular Genetics (IGMM), CNRS UMR5535 and University of Montpellier, 1919 route de Mende, 34293 Montpellier, France, [3]Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, [4]Department of Medical and Molecular Genetics, King's College London, Guy's Hospital, London, SE1 9RT, UK, [5]Servicio de Neonatología, Hospital Sant Joan de Déu, Barcelona, Spain and [6]UMR 3215/Inserm U934, Unité de génétique et biologie du développement, Institut Curie, 26 rue d'Ulm, Paris Cedex 05, France

Received October 7, 2010; Revised November 5, 2010; Accepted November 12, 2010

ABSTRACT

Imprinted retrotransposed genes share a common genomic organization including a promoter-associated differentially methylated region (DMR) and a position within the intron of a multi-exonic 'host' gene. In the mouse, at least one transcript of the host gene is also subject to genomic imprinting. Human retrogene orthologues are imprinted and we reveal that human host genes are not imprinted. This coincides with genomic rearrangements that occurred during primate evolution, which increase the separation between the retrogene DMRs and the host genes. To address the mechanisms governing imprinted retrogene expression, histone modifications were assayed at the DMRs. For the mouse retrogenes, the active mark H3K4me2 was associated with the unmethylated paternal allele, while the methylated maternal allele was enriched in repressive marks including H3K9me3 and H4K20me3. Two human retrogenes showed monoallelic enrichment of active, but not of repressive marks suggesting a partial uncoupling of the relationship between DNA methylation and repressive histone methylation, possibly due to the smaller size and lower CpG density of these DMRs. Finally, we show that the genes immediately flanking the host genes in mouse and human are biallelically expressed in a range of tissues, suggesting that these loci are distinct from large imprinted clusters.

INTRODUCTION

Genomic imprinting is a form of epigenetic gene regulation that results in allelic expression dictated by parental origin (1). Differential DNA methylation is a major component in regulating this process. Discrete differentially methylated *cis*-acting regions, known as imprinting control regions (ICRs), orchestrate the monoallelic expression of numerous genes within imprinted domains and are established while the maternal and paternal genomes are physically separated in their respective germ lines. To date, all imprinted domains are known to contain regions of differential DNA methylation (DMRs) that are deposited in CpG-rich sequences during oogenesis or spermatogenesis by the DNMT3A/DNMT3L *de novo* methyltransferase complex (2–4). A subset of maternally DNA-methylated germline DMRs require the activity of the amine oxidase domain 1 containing histone demethylase AOF1/KDM1. This demethylase is presumably needed to remove any permissive histone H3 lysine 4 (H3K4) methylation present at these CpG islands in the growing oocytes (5). After fertilization, these regions of differential DNA methylation are maintained in somatic

*To whom correspondence should be addressed. Tel: +34 93260 7500 (ext. 3165); Fax: +34 93260 7219; Email: dmonk@iconcologia.net

tissues by DNMT1 (6), and are associated with numerous histone modifications.

It has previously been shown that DMRs have a constitutional histone signature that comprises histone H3 lysine 9 trimethylation (H3K9me3), Histone H4 lysine 20 trimethylation (H4K20me3) and symmetrical histone H2A/H4 arginine 3 dimethylation (H2A/H4R3me2s) on the DNA methylated allele (7,8). This is in contrast to the enrichment of the transcriptionally permissive H3K4me2/3 mark on the unmethylated allele (9). A number of genes that are imprinted solely in the mouse placenta, have been shown recently to require allelic repressive histone modifications at their own DNA-unmethylated promoters to maintain allelic expression (10–13).

Imprinted genes have diverse evolutionary origins. Some imprinted genes are products of retrotransposition from parental genes on the X chromosome. Four imprinted retrogenes in the mouse—*Mcts2*, *Nap1l5*, *U2af1-rs1* (also called *Zrsr1*) and *Inpp5f_v2*—are associated with DMRs at their promoters, and reside in introns of multi-exonic host genes. In all cases, at least one transcript of the host is also subject to imprinting. We have previously shown that the mouse retrogene *Mcts2* influences the choice of polyadenylation (polyA) site for transcripts of the host gene *H13* in an allele-specific manner (14). Expression of *Mcts2* from the paternal allele causes *H13* transcripts to terminate upstream of the retrogene. On the maternal chromosome, *H13* utilizes downstream polyA sites because the *Mcts2* DMR is methylated and the retrogene silenced. A recent transcriptome-wide analysis, using the ultra sensitive RNA-seq technology which is capable of detecting subtle biases in allelic transcription, has suggested that one transcript variant of *Inpp5f*, the host gene of *Inpp5f_v2*, and *Herc3*, the *Nap1l5* host gene, are also subject to isoform-specific allelic expression (15).

To investigate whether the human orthologues of the X-derived imprinted retrogenes influence allele-specific expression of their respective host genes, and whether this influence extends to neighboring genes, we have analyzed the allelic expression of retrogenes, host and flanking genes in humans. The *U2af1-rs1* gene does not have a human counterpart. We find that the human orthologues *INPP5F_V2*, *NAP1L5* and *MCTS2* are paternally expressed in a wide range of fetal tissues, and that their promoters are embedded in maternally DNA-methylated regions. In humans, the host genes are not subject to imprinting, probably due to differing exon/3′-UTR positions in relation to the retrogene integration sites. The genes immediately flanking the host genes are biallelically expressed in both mice and humans, showing retrogene-host pairs do not form parts of larger imprinted clusters. In mice, the allelic chromatin of these DMRs conforms to the constitutional histone modification signature with the repressive modifications H3K9me3, H4K20me3 and H2A/H4R3me2s enriched on the DNA methylated allele, and the permissive modification H3K4me2 enriched on the unmethylated allele. These patterns of histone modifications are not conserved at the human *NAP1L5* and *MCTS2* promoters, correlating with reduced CpG content and CpG island size, which we speculate may influence the recruitment of the histone methyltransferases (HMTs).

MATERIALS AND METHODS

Human tissues

A cohort comprising 65 fetal tissue sets (8–18 weeks) with corresponding maternal blood sample and 96-term placental samples are from the Moore Tissue bank and is described elsewhere (16). An additional 96 human placenta samples were obtained from the Hospital St Joan De Deu collection (Barcelona, Spain). Normal peripheral blood was collected from adult volunteers aged between 19–60-years old. DNA and RNA extraction and cDNA synthesis were carried out as previously described (11). Ethical approval for adult blood and fetal tissue collection was granted by the Hammersmith, Queen Charlotte's and Chelsea and Acton Hospital Research Ethics Committee (Project Registration 2001/6029 and 2001/6028); Collection of the HSJD placental cohort was granted by the ethical committee of Hospital St Joan De Deu Ethics Committee (Study number 35/07).

Cell lines and mouse crosses

Wild-type mouse embryos and placentas were produced by crossing C57BL/6 females with either *Mus musculus molosinus* (JF1) or *Mus musculus castaneus* (C) male mice. RNA and DNA from *Dnmt3l*$^{-/+}$ mice (BxC, C57BL/6 mother and Castaneus father) was isolated and extracted as previously described (2). The E9.5 *Dnmt3l*$^{-/+}$ embryos (BxJ) used to assess *Nap1l5* expression were a kind gift from Dr Kenichiro Hata (NRICHD, Okura, Tokyo, Japan). The human TCL1 and 2 placental trophoblast cell lines were grown in DMEM supplemented with 10% FCS and antibiotics.

Allelic expression analysis

Genotypes of DNA were obtained for exonic SNPs identified in the UCSC browser (NCBI36/hg18, Assembly 2006) by PCR. Sequences were interrogated using Sequencher v4.6 (Gene Codes Corporation, MI, USA) to distinguish informative heterozygote samples. Informative samples were analysed by RT–PCR in corresponding cDNA using, where possible, intron-crossing primers that incorporated the heterozygous SNP in the resulting amplicon (Supplementary Table S1). RT–PCRs were performed using cycle numbers determined to be within the exponential phase of the PCR, which varied for each gene, but was between 32–40 cycles. The RT–PCRs for *HERC3A*, *HERC3C* and both isoforms of *Abcg2* were analyzed by nested RT–PCR, with the first PCR amplified for 25 cycles, with 5 µl of this product used as template for the second round PCR which was limited to 30 cycles.

Real-time qRT–PCR

All PCRs were run in triplicate from the same sample on either an ABI Prism 7700 sequence detector or a 7900 Fast real-time PCR machine (Applied Biosystems) following

the manufacturer's protocol. All primers were optimized using SYBR Green amplification followed by melt curve analysis to ensure that amplicons were free of primer dimer products. Thermal cycling parameters included Taq polymerase activation at 95C for 10 min for one cycle, repetitive denaturation at 95°C for 15 s, and annealing at 60°C for 1 min for 40 cycles. All resulting triplicate cycle threshold (C_t) values had to be within one C_t of each other. The quantitative values for each triplicate were determined as a ratio with the level of *Gapdh*, measured in the same sample, with the mean providing relative expression values.

Analysis of allelic DNA methylation

Approximately 1 μg DNA was subjected to sodium bisulphite treatment and purified using the EZ GOLD methylation kit (ZYMO, Orange, CA, USA). Bisulphite specific primers for each region were used with Hotstar Taq polymerase (Qiagen, West Sussex, UK) at 45 cycles and the resulting PCR product cloned into pGEM-T Easy vector (Promega) for subsequent sequencing. MeDIP was performed on 6 μg sonicated (Diagenode Bioruptor) genomic DNA with an average size of 150 bp. Samples were denatured and incubated with a monoclonal antibody against 5-methylcytidine (Eurogentec). Immunoprecipitated DNA was then isolated using IgG Dynabeads (Dynal Biotech), digested with proteinase K and phenol–chloroform extraction was followed by ethanol precipitation. MeDIP enrichment was verified by duplexed PCR for the methylated *SERPIN* B5 promoter and the unmethylated *UBE2B* promoter. Southern blotting was performed following standard protocols using methylation-sensitive restriction enzymes. Digested DNA was subjected to agarose gel electrophoresis and transferred to Hybond N+ membrane (Amersham). Radio-labeled PCR product probes were hybridized over-night at 65°C, and subsequently washed in increasing stringency SSC/0.1% SDS washes. PCR primers used to generate probes are listed in Supplementary Table S1.

Chromatin immunoprecipitation

Two adult leukocyte samples and the TCL1 and TCL2 cell lines were used in addition to E18.5 mouse embryos for chromatin immunoprecipitation (ChIP). ChIP was carried out as previously described (11,13) using the following Upstate Biotechnology antisera directed against H3K4me2 (07-030), H3K9me2 (07-441), H3K9me3 (060904589), H3K9ac (07-352), H3K27me3 (07-449), H4K20me3 (07-463) (Upstate Biotechnology) and H2A/H4R3me2s (Abcam ab5823 97454/520317). ChIPed DNA was subjected to allele-specific PCR. Polymorphisms within 1 kb of the CpG island were identified by interrogating SNP databases or genomic sequencing (see Supplementary Table S1 for primer sequences and location). Only ChIP sample sets that showed enrichment for additional ICRs were used in the analysis.

Precipitation levels in the ChIP samples were determined by real-time PCR amplification, using SYBR Green PCR kit (Applied Biosystems). Each PCR was run in triplicate and results are presented as fold enrichment (comparison to mock) and normalized to the level of precipitation at the *SNURF*-ICR, a control for both active and repressive histone modifications located on human chromosome 15.

RESULTS

MCTS2 does not influence allelic expression of *HM13*

The *H13* gene on mouse chromosome 2 is known to generate at least five transcripts, all originating from a single promoter, but differing in polyA site usage (14). The utilization of these alternative polyA sites is influenced by the paternally expressed *Mcts2* imprinted retrogene and the CpG island that comprises its DMR. The short *H13d* and *e* transcripts are paternally expressed, whereas the *H13a*, *b* and *c* transcripts, that extend through *Mcts2* and the DMR to the canonical polyA site are maternally expressed (Figure 1A). To assess allelic expression in humans we identified transcribed SNPs unique to each isoform, and allele-specific assays were carried out in a selection of human first trimester fetal tissues and term placentas. The human *MCTS2* gene (also known as *PSIMCT-1*, MCTS1-pseudogene) is imprinted in a variety of fetal tissues (Figure 1B and Supplementary Table S2). *MCTS2* differs from its mouse orthologue as it can splice into the last eight exons of the *HM13* host gene (Figure 1B). Sequence analyses revealed that in addition to the highly conserved MCTS2 open reading frame, this RNA has the potential to be bi-cistronic, encoding for a chimeric protein lacking the first 151 amino acids of HM13, but sharing 243 amino acids in the C-terminus.

The *MCTS2* promoter is embedded within a DMR (Figure 1B and Supplementary Figure S1A), while the *HM13* host gene originates from a CpG island that is unmethylated in all the tissues analysed. Alignment of human expressed sequence tags (ESTs) revealed a number of transcripts. The expression of the human short *HM13D* isoform (Genbank NM_178982), which terminates prior to the transcriptional start site (TSS) of *MCTS2*, and the full-length transcript *HM13C* (Genbank NM_030789) are biallelic in all fetal tissues analysed (Figure 1B and Supplementary Table S2).

The *NAP1L5* promoter is not within a CpG island, but is a DMR

The paternally expressed *Nap1l5* retrogene was first identified in a genome-wide screen for differential DNA methylation, and was reported to be predominantly paternally expressed in mouse brain (17). The host gene, *Herc3*, gives rise to a number of transcript isoforms. At least two short isoforms (*Herc3b* and *c*) are expressed from the paternal allele in mouse brain [A.J. Wood and R.J. Oakey, unpublished data, (15)]. The full length *Herc3a* transcript has a maternal expression bias, similar to the full length *H13* isoforms (14). We detect paternal expression of *NAP1L5* in all fetal tissues analyzed. The human *HERC3* gene also contains one long (*HERC3A*, Genbank NM_014606) and two short isoforms (*HERC3B*, Genbank BC038960; *HERC3C*, Genbank AK296397).

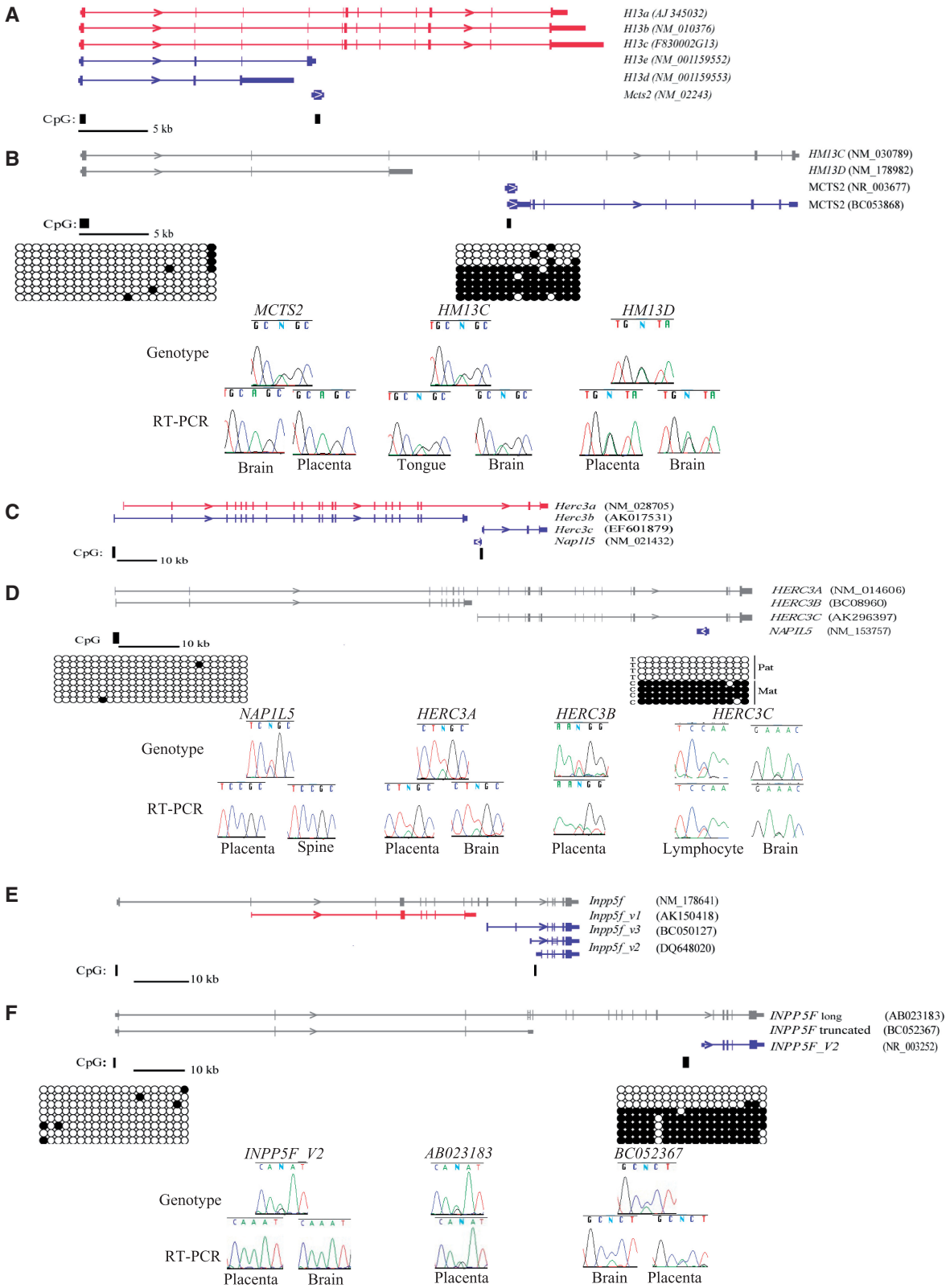

Figure 1. (**A**) Map of the *H13* locus located on mouse chromosome 2, showing the location of the various imprinted transcripts and CpG islands (red transcripts are maternally expressed, blue are paternally expressed and grey are expressed from both parental alleles. Arrows represent direction of transcription). (**B**) Schematic of the human *HM13* gene on chromosome 20, showing the distribution of exons and insertion of *MCTS2* into intron 4. The methylation status of the *HM13* promoter CpG island and *MCTS2* CpG island were examined by bisulphite PCR. Each circle represents a single CpG dinucleotide and the strand. Filled circle, a methylated cytosine; open circle, unmethylated cytosine. The sequence traces show allelic expression for *MCTS2* and *HM13* isoforms (for clarity only sequence traces for *MCTS2* BC053868 are shown). (**C**) A map of the *Herc3* domain on mouse chromosome 6. (**D**) The human *NAP1L5* gene and the insertion into intron 22 of *HERC3*. (**E**) A schematic map of the *Inpp5f* gene on mouse chromosome 7, and (**F**) the orthologous region on human chromosome 10.

Our allele-specific assays are suggestive of biallelic expression of all isoforms in all tissues including brain (Figure 1C and D, and Supplementary Table S2), although our methodology is not directly comparable with RNA-seq, which can detect subtle expression biases. The *HERC3* transcripts initiate from an unmethylated CpG island, whereas the *NAP1L5* promoter is DNA methylated on the maternal allele, even though the region is not statistically a CpG island in humans (Figure 1D).

INPP5F_V2 is imprinted in numerous human tissues

We previously identified a neural-specific, paternally expressed *Inpp5f_v2* transcript using expression microarrays (18). Like the other murine imprinted retrogene loci, the host gene *Inpp5f* exhibits isoform-specific expression, with a maternal expression bias in a truncated shorter transcript (Genbank AK039468) (15). In addition, another transcript, *Inppf5_v3*, arising from a different promoter, is paternally expressed in mouse brain (19,20). The genomic organization of the human locus resembles that of the mouse, however, there is no evidence for a human *Inpp5f_v3* orthologue (Figure 1E and F). The promoter of the human *INPPF5_V2* transcript is embedded within a maternally DNA-methylated DMR (Figure 1F and Supplementary Figure S1B), resulting in paternal expression in a wide range of tissues (Supplementary Table S2). The full-length host gene transcripts (Genbank AB023183), and the truncated isoform (Genbank BC052367), originate from an unmethylated CpG island and are biallelically expressed.

Retrogenes-host pairs do not form part of larger imprinting clusters

In order to determine the boundaries of imprinting at each retrogene-host locus, we investigated the allele-specific expression of the genes flanking the host genes in both mice and humans. We assessed expression in various embryonic tissues and placenta using allele-specific assays between crosses of mouse strains C57BL/6 (B) x Mus musculus castaneus (C). The genes immediately adjacent to *Mcts2/H13*, *Id1* and *Remi*, are biallelically expressed in all tissues at embryonic day E18.5, as are *111007A13RIK* and *Bag3* flanking *Inpp5f_v2/Inpp5f*. The *Fam13a* gene, telomeric to *Nap1l5/Herc3* is expressed from both alleles, while *Abcg2*, centromeric to *Nap1l5/Herc3*, is monoallelically expressed in the placenta. This reflects a bias in expression from the C57BL/6 allele and the gene is therefore an expressed quantitative trait locus (eQTL) and not imprinted (Figure 2A and Supplementary Figure S2). This conclusion is supported by the persistence of *Abcg2* monoallelic expression in *Dnmt3l*$^{-/+}$ placental trophoblast, despite the loss of imprinting of *Nap1l5* (Figure 2B).

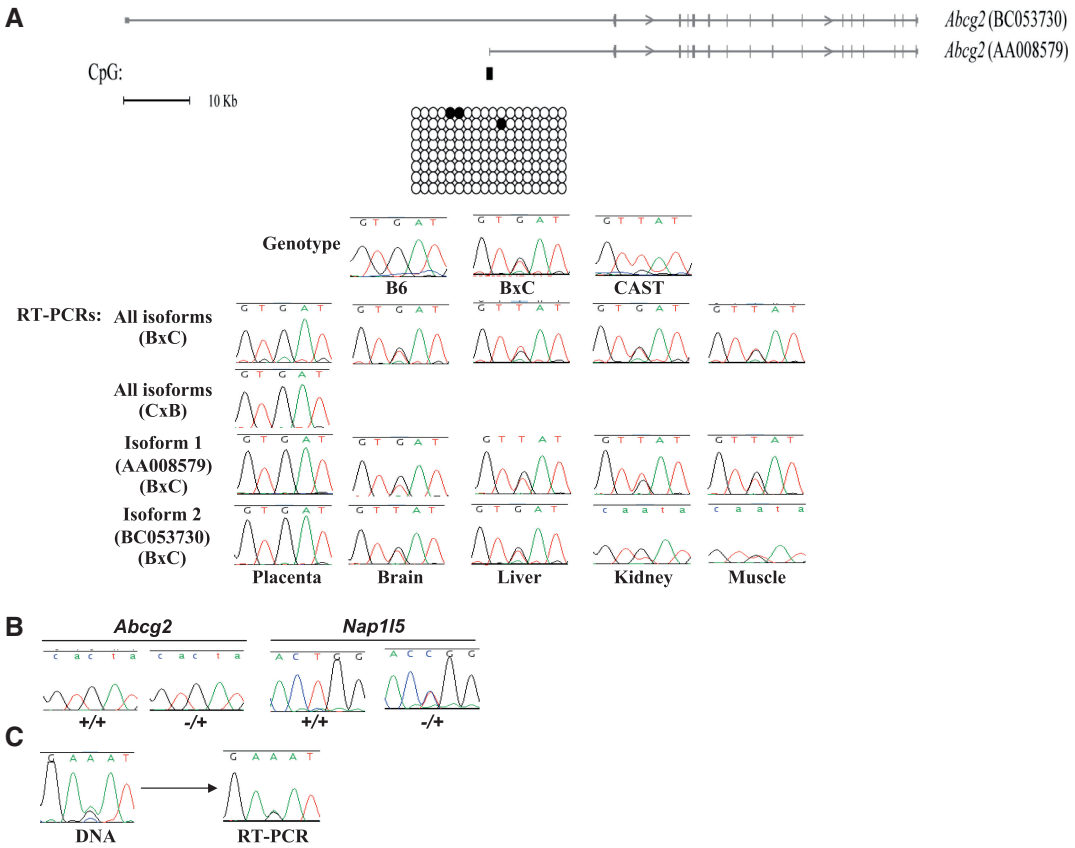

Figure 2. (**A**) A schematic map of the *Abcg2* gene, with the location of the alternative promoter regions. The methylation status of the CpG island associated with isoform 1 was examined in placenta-derived DNA. The allelic expression of *Abcg2* is assessed in various fetal tissues in reciprocal mouse crosses. (**B**) The allelic expression of *Abcg2* and *Nap1l5* in placental trophoblasts from *Dnmt3l*$^{-/+}$ mice. (**C**) The allelic expression of the *ABCG2* gene in human term placenta.

To assess the allelic origin of expression for the orthologous flanking genes in the human, we assessed the expression of *REMI*, *ID1*, *BAG3*, *CORF119* and *FAM13A*/*FAM13AOS*. These genes are expressed biallelically in all fetal tissues (Supplementary Table S3 and Figure S3). In higher primates, including rhesus monkey, orangutan, chimps and humans, the organization of the genes centromeric to *NAP1L5*/*HERC3* is different to that of mouse and rat, resulting in the gene *PIGY* being immediately centromeric to *NAP1L5*/*HERC3*, and *ABCG2* ~300 kb away. In all human fetal tissues analyzed, including placenta, both *PIGY* and *ABCG2* are biallelic (Figure 2C and Supplementary Figure S3). These finding strongly suggest that imprinted retrogene-host pairs do not form part of larger imprinting clusters.

Histone modification and allelic repression at imprinted retrogene DMRs

In mouse the imprinted expression of *Nap1l5* and *Inpp5f_v2* is restricted to the brain (17–19), whereas in humans, imprinted expression is observed in a wider variety of fetal tissues (Figure 1, Supplementary Table S1 and Figure S4). To date all germline DMRs are differentially methylated in all somatic tissues, indicating that DNA methylation on its own is not responsible for tissues-specific differences in expression observed at imprinted loci. To investigate whether the discrepancy in expression profiles we observe between species could be attributed to histone modifications, we analyzed the allelic enrichment of both permissive and repressive histone modifications. Our analysis focused on modifications on histone H3 and H4, including acetylation of lysine-9 (H3K9ac) and H3K4me2 as markers of active chromatin; and the repressive marks of H3K9me3 and H3K27me3 of histone H3, along with the histone H4 modifications H4K20me3 and H2A/H4R3me2s.

ChIP was performed on native chromatin from brain and decapitated embryos for both BxC and BxJF1 crosses. We ascertained allelic enrichment using polymorphisms mapping within the DMRs of *Nap1l5, Inpp5f_v2* and *Mcts2*. The active modification H3K4me2 was strongly enriched specifically on the unmethylated paternal allele for all three DMRs in both brain and embryo, while most H3K9ac precipitation was predominantly in brain (Figure 3), the tissue in which these genes are expressed. The same regions showed precipitation of the repressive marks H3K9me3, H4K20me3 and H2A/H4R3me2s on

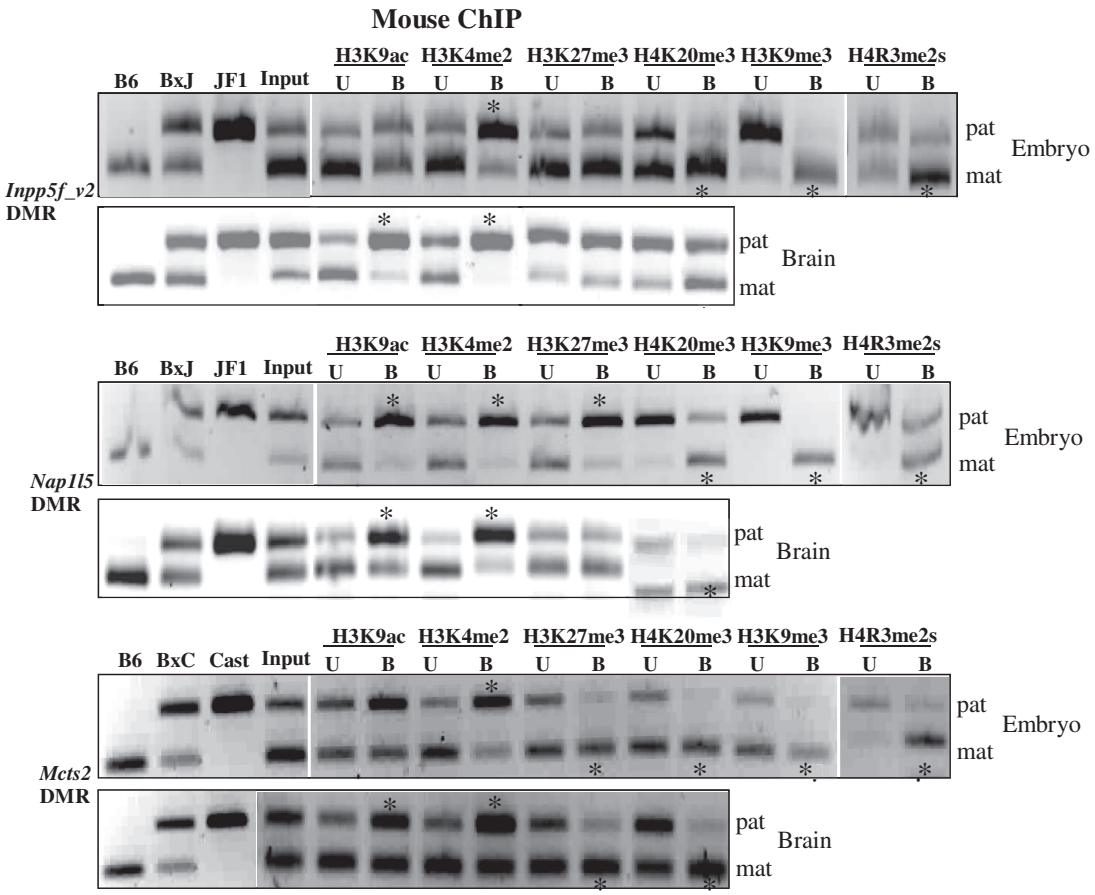

Figure 3. (**A**) The allelic precipitation of the mouse retrogene DMRs in embryo and brain tissues. Native ChIP followed by PCR and restriction digest-mediated allelic discrimination of the input, antibody bound (**B**) and unbound (U) chromatin fractions on BxJ embryos and brains for *Nap1l5* and *Inpp5f_v2*, and on BxC embryos and brains for *Mcts2*. The asterisks represent a relative allelic enrichment of >3-fold compared to the unbound fraction.

the DNA methylated maternal allele (Figure 3). The repressive H3K27me3 mark showed different profiles for the three mouse DMRs. This modification did not show allelic enrichment at the *Inpp5f_v2* DMR in either brain or embryo, but was precipitated on the DNA methylated maternal allele at the *Mcts2* DMR. At the *Nap1l5* DMR, we observed that H3K27me3 was precipitated on the unmethylated paternal allele in embryo but not in brain (Figure 3). This pattern of enrichment is reminiscent of the monoallelic bivalent chromatin domain reported at the *Grb10*/*GRB10* gene (21,22). This monoallelic bivalent conformation is not detected at the *Nap1l5* DMR in brain, suggesting that the removal of H3K27me3 is concomitant with the paternal expression observed for *Nap1l5* in mouse brain.

Extensive genotyping of the human *NAP1L5*, *MCTS2* and *INPP5F_V2* DMRs revealed that SNPs in these regulatory regions are rare. However, we were able to identify heterozygous samples that allowed us to discriminate between alleles. The SNP rs2972011 is located ~200 bp from the TSS of *NAP1L5*, whereas rs7907781 and rs1115713 are ~600 bp and ~50 bp from the TSS of *INPP5F_V2* and *MCTS2*, respectively. To ensure that these SNPs mapped within the DMRs, we performed DNA methylation immunoprecipitation (meDIP) using antisera directed against 5-methylcytosine. This was due to the difficultly in amplifying bisulphite converted DNA in the vicinity of SNPs rs7907781 and rs1115713. For all three regions we observed monoallelic enrichment in heterozygous placental DNA samples, and where informative, the DNA methylation was detected on the maternal allele (Figure 4A). Using these same amplification conditions, we performed ChIP on native chromatin isolated from adult peripheral blood leukocytes and from two human placental cell lines, TCL1 and TCL2 (23) for the *NAP1L5* and *MCTS2* DMRs. Unfortunately, no heterozygous cell lines could be found that were informative for *INPP5F_V2*, despite genotyping of over 140 leukocyte samples and normal tissue cell lines. Similar to the mouse, we observe strong monoallelic enrichment for H3K4me2 at the *NAP1L5* and *MCTS2* DMRs, but since no parental DNA samples were available, allelic origin could not be assigned. Unexpectedly, we did not observe allelic precipitation for any of the repressive histone marks at these DMRs, despite strong allelic enrichment at the *SNURF*/*SNRPN*, *H19* and *MEST* DMRs (Figure 4B; data not shown). To confirm that the histone modifications were present at the *NAP1L5* and *MCTS2* DMRs, we performed quantitative ChIP analysis on the placental cell line TCL1 (Figure 4C). The precipitation values obtained were normalized to those for the *SNURF*/*SNRPN* DMR, which revealed that H3K4me2 is more abundant at the *MCTS2* and *NAP1L5* promoters, whereas, the repressive histone modifications were precipitated several fold less. Interrogation of human histone maps [http://dir.nhlbi.nih.gov/papers/lmi/epigenomes/hgtcell.aspx, and (24)] confirmed the absence of significant enrichment for these repressive marks at 1–2 nucleosomes resolution

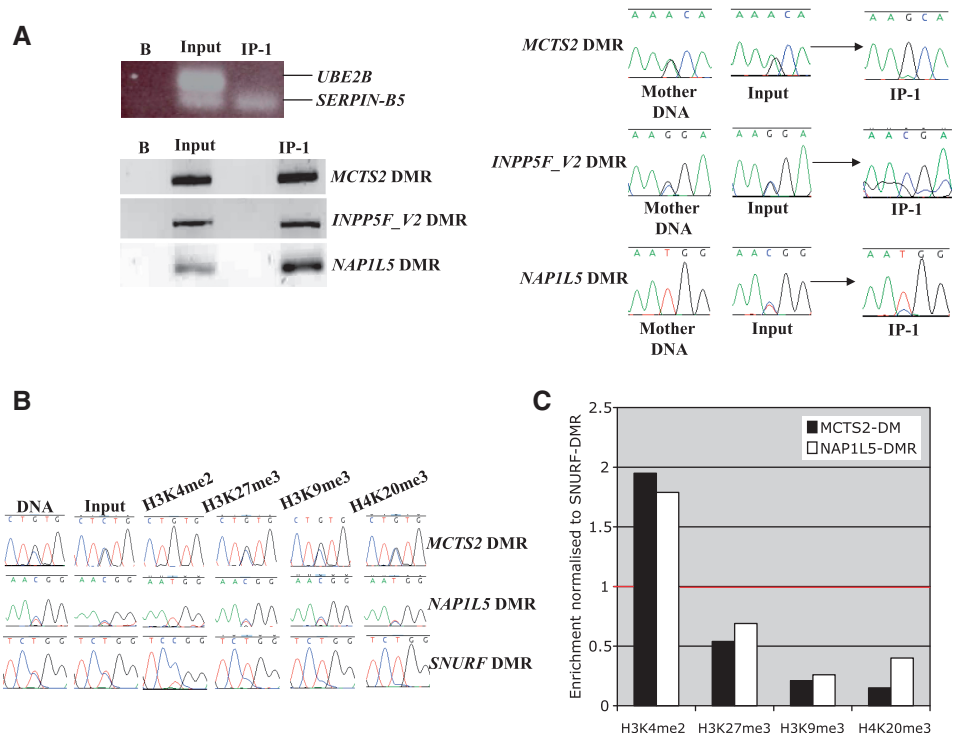

Figure 4. (**A**) Methylation-immunoprecipitation was performed on placental DNA using anti 5mC antibody. The efficiency of IP was assessed by PCR specific for the methylated *SERPIN-B5* promoter and the unmethylated *UBE2B* promoter. The precipitations were subsequent used to assess the methylation at the human *MCTS2*, *INPP5F_V2* and *NAP1L5* DMRs. (**B**) Using the same PCR primer combinations, allelic-ChIP was performed on human placental cell lines (for clarity, only ChIP-bound fractions are shown). (**C**) qPCR on ChIP-bound material from the TCL1 cell immunoprecipitations. Levels of precipitation are compared to the *SNURF* DMR (red line, equal to one).

(the approximate size of the *MCTS2* and *NAP1L5* DMRs), despite strong H3K4me2 enrichment. Together, these data suggest that these repressive marks are not present, or very low, in these two regions.

DISCUSSION

In this study, we have shown the paternal allele-specific expression of *INPP5F_V2*, *MCTS2* and *NAP1L5* in a wide range of human tissues. This is in contrast to the mouse, where *Inpp5f_v2* and *Nap1l5* show spatial expression restricted to brain. We show that the TSSs for these three genes are embedded in regions of differential DNA methylation, similar to their mouse orthologues.

In the mouse, *Mcts2* and *Nap1l5*, like *Inpp5f_v2* are organized such that the imprinted 'intronic retrogene' and its DMR/promoter reside in the intron of another gene known as the 'host'. In each case the retrogene originated from an ancestoral gene on the X chromosome some time in early eutheraian evolution, since *Inppf5_v2*, *Mcts2* and *Nap1l5* are absent in marsupials (19). Retrotransposition has also been linked to the imprinting of the *RB1* gene, however this event occurred much later in mammalian evolution as only humans, chimpanzee and rhesus monkeys, but not mice and rat have the processed *RB1/KIAA0649* pseudogene (25). A recent high-resolution analysis of parent-of-origin allelic expression in mouse brain has revealed that the host genes *Herc3* and *Inpp5f* show evidence for allele-specific alternative polyA choice similar to that of *H13* [A.J. Wood and R.J. Oakey, unpublished data and (14,15)].

To assess whether the human orthologues *MCTS2*, *NAP1L5* and *INPP5F_V2* are also associated with imprinted host genes, we analyzed the allelic expression of the host genes in a wide range of fetal tissues and term placentas. We observe in all cases, that the host transcripts are biallelically expressed. The genomic location of the host gene exons are different. In the mouse, the polyA signal for the paternally expressed *H13d* isoform maps to within 100 bp of the *Mcts2* DMR, whereas these two features are separated by >7 kb in humans. Genomic rearrangements are more pronounced at the *Nap1l5/NAP1L5* locus, where *Herc3b* and *Nap1l5* are separated by 1.5 kb in the mouse, and by more that 39 kb in humans. In addition, the mouse *Herc3c* isoform initiates from within the *Nap1l5* DMR, whereas the human promoter for this isoform is adjacent to the *HERC3B* polyA (Figure 1D). These differences in exon distribution could explain the lack of host gene imprinting observed in humans. We have previously proposed that transcriptional interference may be involved in alternative polyA choice at *H13*, due to the high expression of *Mcts2* in brain directly inhibiting the transcription of *H13* on the paternal allele in co-expressing cells. Alternative models, including the recruitment of methylation-sensitive polyA factors, or the association with the CTCF/cohesin boundary complex to the unmethylated paternal allele of the *Mcts2*, could result in similar allelic-termination of the host genes on the maternal allele (14). For any of the models, the imprinting of host gene transcripts may require a close physical proximity of the host gene exons with the retrogene, which is not observed in humans.

Genes flanking the retrogene-host pairs are biallelically expressed

Many imprinted genes are clustered in the genome, such that their expression is influenced by shared control elements, such as DMRs. The imprinted retrogene-host pairs are located outside of characterized clusters. To confirm this, we assayed flanking genes for expression status and found all to be biallelic, with the exception of *Abcg2* which is monoallelically expressed in the placentae from reciprocal BxC and CxB sub-species intercrosses, but not subject to imprinting.

It has recently become evident that non-imprinted monoallelic expression can result from SNP-associated DNA methylation (26). To assess whether the *Abcg2* eQTL is due to monoallelic DNA methylation similar to that described in humans, we showed that the promoter CpG island of *Abcg2* is fully unmethylated (Figure 2 and data not shown) suggesting that another mechanism is regulating the allelic expression. Recently, Brideau *et al.* (27) reported that the *AK006067* transcript next to the imprinted *Rasgrf1* gene is expressed >100-fold higher from the C57BL/6 allele than the PWK allele. Both our finding of the *Abcg2* eQTL and the observations of Brideau *et al.* (27), emphasize the necessity to study reciprocal F1 crosses.

Histone modification signatures at imprinted retrogene DMRs

Recent studies have suggested that there is a link between DNA and histone methylation at imprinted DMRs. A comprehensive analysis of allelic histone modifications in *Dnmt3l*$^{-/+}$ conceptuses revealed that without oocyte-derived DNA-methylation imprints, there is a dramatic effect on the presence of repressive histone modifications, with maternally DNA methylated DMRs adopting a paternal epigenotype (7). We explored whether the chromatin at *Nap1l5*, *Mcts2* and *Inpp5f_v2* is associated with the plethora of histone modifications known to be enriched at DMRs. In agreement with our previous observations, we found that the DNA-methylated alleles of each DMR are enriched for the repressive histone marks H3K9me3, H4K20me3 and H2A/H4R3me2s. Unlike these three modifications, H3K27me3 was not consistently associated with the DNA-methylated allele, which agrees with earlier chromatin studies on ICRs. At the *Nap1l5* DMR, we confirm the previous observation of bivalent chromatin in mouse embryos (7), with H3K4me2 and H3K27me3 both enriched on the paternal allele. This monoallelic bivalent domain behaves in a similar fashion to the recently described monoallelic bivalent chromatin at the *Grb10* DMR (21). Like their non-imprinted bivalent counterparts, these genes are associated with 'poised' lineage-specific transcription in mouse and human ES cells (28). In brain, we observe absence of allelic enrichment for H3K27me3, and this correlates with acquired expression of *Nap1l5*, a mechanism comparable to what we have reported previously at the *Grb10* (21,22).

These observations suggest that monoallelic bivalent chromatin could be a common mechanism conferring brain-specific imprinted gene expression.

To confirm the conservation of these histone modifications at the regions we identified as DMRs in humans, we performed ChIP on leukocytes and placenta cell lines. To our surprise we did not find significant enrichment of the repressive histone marks H3K9me3 and H4K20me3 on the DNA-methylated allele. This uncoupled action of the DNA-methylation machinery and the H3K9me3 and H4K20me3 HMTs may be partially explained by the progressive decrease in CpG island size and CpG density at the *NAP1L5* and *MCTS2* DMRs compared to other imprinted DMRs in the genome (Supplementary Table S4). Throughout mammalian evolution, the *NAP1L5* and *MCTS2* DMRs have lost approximately half of their CpG dinucleotides compared to mouse. Both DMRs have diminished in size, with a reduction from >420 bp to 216 bp for *MCTS2*. The human *NAP1L5* DMR fails to reach standard CpG island criteria (GC content >50%; Obs CpG/Exp CpG >0.6; min length 200 bp). The loss in CpG island size at the *NAP1L5* DMR is due to a combination of CpG deamination and the integration of numerous CpG low-density repeat elements, including DNA-MER115, LINE-1 and low-complexity CT-repeats immediately downstream of the transcription start site. The human and mouse genomes have recently been shown to contain a similar number of CpG islands (29). Our observation of an evolutionary loss of CpG density at the *NAP1L5* and *MCTS2* DMRs goes against the general trend. Additionally, for many imprinted DMRs analyzed in humans (*Kv*DMR1, *GRB10*, *MEST*, *ZAC1*, *NDN*, *GNAS EX1A*, *GNAS XL*, *PEG3*, *PEG10*, *NNAT* and *IGF2R*/*AIR*), the size of the CpG islands comprising the DMRs are all larger in humans than in mouse (Supplementary Table S4). The low-CpG density within the promoters of *NAP1L5* and *MCTS2* may mean that these discrete DMRs go unrecognized by the non-histone proteins including the HMTs for H4K20me3 and H3K9me3, respectively (8,12). Overall these observations are in agreement with the recent studies suggesting that DNA methylation at functional imprints require DNA methylation before the acquisition of repressive histone methylation (7), and that deficiencies in repressive histone marks do not have a direct role in the regulation of DNA methylation at ICRs (12,13,30).

CONCLUSIONS

In summary, we have shown that the human orthologues of mouse imprinted retrogenes are paternally expressed in a wide range of fetal tissues. In mice, the host genes are subject to alternative polyadenylation, presumably as a consequence of retrogene integrations that acquired imprinting in proximity to weak polyA signals. In humans, we show that retrogene promoters are subject to allele-specific CpG methylation, but internal polyA sites of host genes are situated further upstream of the DMRs and thereby escape their influence. In mice, these DMRs are associated with allelic repressive histone modifications. At the mouse *Nap1l5* promoter, monoallelic bivalent chromatin i.e. the enrichment of both H3K4me2 and H3K27me3 on the same allele, is associated with the unmethylated paternal allele. In humans, an evolutionary deterioration in CpG island size correlates with a lack of allelic H3K9me3 and H4K20me3 precipitation at *NAP1L5* and *MCTS2* DMRs indicating imprinted gene expression in the absence of the an ICR-specific histone signature.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank R. Schulz for critical reading of the article and C. Landles for technical assistance with Southern blotting. We thank K. Hata for *Dnmt3l*$^{-/+}$ embryos on the BxJ background and Dr M. Sullivan for the TCL cell lines.

FUNDING

Spanish Ministerio de Educacion y Ciencia (grant number SAF2008-1578 to D.M.); Association Pour La Recherche sur le Cancer (grant number 4980; 2010 to P.A.); La Ligue contre le Cancer (grant 'subvention comite' Herault: 2010 to P.A.); CNRS 'Projects for International Scientific Cooperation' (to P.A. and D.M.); Agence National de la Recherche; 'Institut National du Cancer'; Association for International Cancer Research; 'Ligue Contre le Cancer' (to R.F.); Medical Research Council (to G.E.M.); SPARKs; Wellbeing for Women (to G.E.M.); The Wellcome Trust (to R.J.O., J.M.F. and M.C.); Ramon y Cajal research fellowship (to D.M.); FPU studentship (to A.G.A.). Funding for open access charge: Internal funding.

Conflict of interest statement. None declared.

REFERENCES

1. Verona,R.I., Mann,M.R. and Bartolomei,M.S. (2003) Genomic imprinting: intricacies of epigenetic regulation in clusters. *Ann. Rev. Cell Dev. Biol.*, **19**, 237–259.
2. Bourc'his,D., Xu,G.L., Lin,C.S., Bollman,B. and Bestor,T.H. (2003) Dnmt3L and the establishment of maternal genomic imprints. *Science*, **21**, 2536–2539.
3. Kaneda,M., Okano,M., Hata,K., Sado,T., Tsujimoto,N., Li,E. and Sasaki,H. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature*, **24**, 900–903.
4. Kato,Y., Kaneda,M., Hata,K., Kumaki,K., Hisano,M., Kohara,Y., Okano,M., Li,E., Nozaki,M. and Sasaki,H. (2007) Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.*, **16**, 2272–2280.
5. Ciccone,D.N., Su,H., Hevi,S., Gay,F., Lei,H., Bajko,J., Xu,G., Li,E. and Chen,T. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature*, **461**, 415–418.
6. Cirio,M.C., Ratnam,S., Ding,F., Reinhart,B., Navara,C. and Chaillet,J.R. (2008) Preimplantation expression of the somatic

form of Dnmt1 suggests a role in the inheritance of genomic imprints. *BMC Dev. Biol.*, **8**, 9.
7. Henckel,A., Nakabayashi,K., Sanz,L.A., Feil,R., Hata,K. and Arnaud,P. (2009) Histone methylation is mechanistically linked to DNA methylation at imprinting control regions in mammals. *Hum. Mol. Genet.*, **18**, 3375–3383.
8. Pannetier,M., Julien,E., Schotta,G., Tardat,M., Sardet,C., Jenuwein,T. and Feil,R. (2008) PR-SET7 and SUV4-20H regulate H4 lysine-20 methylation at imprinting control regions in the mouse. *EMBO Reports*, **9**, 998–1005.
9. Fournier,C., Goto,Y., Ballestar,E., Delaval,K., Hever,A.M., Esteller,M. and Feil,R. (2002) Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes. *EMBO J.*, **21**, 6560–6570.
10. Umlauf,D., Goto,Y., Cao,R., Cerqueira,F., Wagschal,A., Zhang,Y. and Feil,R. (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nat. Genet.*, **36**, 1296–1300.
11. Monk,D., Arnaud,P., Apostolidou,S., Hills,F.A., Kelsey,G., Stanier,P., Feil,R. and Moore,G.E. (2006) Limited evolutionary conservation of imprinting in the human placenta. *Proc. Natl Acad. Sci. USA*, **103**, 6623–6628.
12. Wagschal,A., Sutherland,H.G., Woodfine,K., Henckel,A., Chebli,K., Schulz,R., Oakey,R.J., Bickmore,W.A. and Feil,R. (2007) G9a histone methyltransferase contributes to imprinting in the mouse placenta. *Mol. Cell. Biol.*, **28**, 1104–1113.
13. Monk,D., Wagschal,A., Arnaud,P., Müller,P.S., Parker-Katiraee,L., Bourc'his,D., Scherer,S.W., Feil,R., Stanier,P. and Moore,G.E. (2008) Comparative analysis of human chromosome 7q21 and mouse proximal chromosome 6 reveals a placental-specific imprinted gene, TFPI2/Tfpi2, which requires EHMT2 and EED for allelic-silencing. *Genome Res.*, **18**, 1270–1281.
14. Wood,A.J., Schulz,R., Woodfine,K., Koltowska,K., Beechey,C.V., Peters,J., Bourc'his,D. and Oakey,R.J. (2008) Regulation of alternative polyadenylation by genomic imprinting. *Genes Dev.*, **22**, 1141–1146.
15. Gregg,C., Zhang,J., Weissbourd,B., Luo,S., Schroth,G.P., Haig,D. and Dulac,C. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, **329**, 643–648.
16. Apostolidou,S., Abu-Amero,S., O'Donoghue,K., Frost,J., Olafsdottir,O., Chavele,K.M., Whittaker,J.C., Loughna,P., Stanier,P. and Moore,G.E. (2007) Elevated placental expression of the imprinted PHLDA2 gene is associated with low birth weight. *J. Mol. Med.*, **85**, 379–387.
17. Smith,R.J., Dean,W., Konfortova,G. and Kelsey,G. (2003) Identification of novel imprinted genes in a genome-wide screen for maternal methylation. *Genome Res.*, **13**, 558–569.
18. Choi,J.D., Underkoffler,L.A., Wood,A.J., Collins,J.N., Williams,P.T., Golden,J.A., Schuster,E.F. Jr, Loomes,K.M. and Oakey,R.J. (2005) A novel variant of Inpp5f is imprinted in brain, and its expression is correlated with differential methylation of an internal CpG island. *Mol. Cell. Biol.*, **25**, 5514–5522.
19. Wood,A.J., Roberts,R.G., Monk,D., Moore,G.E., Schulz,R. and Oakey,R.J. (2007) A screen for retrotransposed imprinted genes reveals an association between X chromosome homology and maternal germ-line methylation. *PLoS Genet.*, **3**, e20.
20. Wood,A.J., Bourc'his,D., Bestor,T.H. and Oakey,R.J. (2007) Allele-specific demethylation at an imprinted mammalian promoter. *Nucleic Acids Res.*, **35**, 7031–7039.
21. Sanz,L.A., Chamberlain,S., Sabourin,J.C., Henckel,A., Magnuson,T., Hugnot,J.P., Feil,R. and Arnaud,P. (2008) A mono-allelic bivalent chromatin domain controls tissue-specific imprinting at Grb10. *EMBO J.*, **227**, 2523–2532.
22. Monk,D., Arnaud,P., Frost,J., Hills,F.A., Stanier,P., Feil,R. and Moore,G.E. (2009) Reciprocal imprinting of human GRB10 in placental trophoblast and brain: evolutionary conservation of reversed allelic expression. *Hum. Mol. Genet.*, **18**, 3066–3074.
23. Takeda,S. and Nakano,H. (2003) TNFalpha-induced apoptosis and integrin switching in human extravillous trophoblast cell line. *Biol. Reprod.*, **68**, 1771–1778.
24. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
25. Kanber,D., Berulava,T., Ammerpohl,O., Mitter,D., Richter,J., Siebert,R., Horsthemke,B., Lohman,D. and Buiting,K. (2009) The human retinoblastoma gene is imprinted. *PLoS Genet.*, **12**, e1000790.
26. Kerkel,K., Spadola,A., Yuan,E., Kosek,J., Jiang,L., Hod,E., Li,K., Murty,V.V., Schupf,N., Vilain,E. *et al.* (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat. Genet.*, **40**, 904–908.
27. Brideau,C.M., Eilertson,K.E., Hagarman,J.A., Bustamante,C.D. and Soloway,P.D. (2010) Successful computational prediction of novel imprinted genes from epigenomic features. *Mol. Cell. Biol.*, **30**, 3357–3370.
28. Bernstein,B.E., Mikkelsen,T.S., Xie,X., Kamal,M., Huebert,D.J., Cuff,J., Fry,B., Meissner,A., Wernig,M., Plath,K. *et al.* (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
29. Illingworth,R.S., Gruenewald-Schneider,U., Webb,S., Kerr,A.R., James,K.D., Turner,D.J., Smith,C., Harrison,D.J., Andrews,R. and Bird,A. (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet.*, **6**, e1001134.
30. Mager,J., Montgomery,N.D., de Villena,F.P. and Magnuson,T. (2003) Genome imprinting regulated by the mouse Polycomb group protein Eed. *Nat. Genet.*, **33**, 502–507.